

⑨ 日本国特許庁 (JP)

⑩ 特許出願公開

⑫ 公開特許公報 (A)

昭59—157699

⑪ Int. Cl.³
G 10 L 1/00

識別記号

庁内整理番号
R 7350—5D

⑬ 公開 昭和59年(1984)9月7日

発明の数 1
審査請求 未請求

(全 6 頁)

⑭ 音声登録方式

① 特 願 昭58—31550

② 出 願 昭58(1983)2月25日

③ 発 明 者 田中厚夫

大阪市阿倍野区長池町22番22号
シャープ株式会社内

④ 発 明 者 上田徹

大阪市阿倍野区長池町22番22号
シャープ株式会社内

⑤ 出 願 人 電子計算機基本技術研究組合

東京都港区三田1丁目4番28号

⑥ 代 理 人 弁理士 根土愛彦 外2名

明 細 書

1. 発明の名称

音声登録方式

2. 特許請求の範囲

1. 音声入力実行時に得られた音節の識別結果の
正誤を指示し、上記音節の識別結果及び正誤の指示にもとず
いて各音節の出現頻度及び誤り度数を求め、上記求められた各音節の出現頻度及び誤り度
数にもとずいて登録あるいは再登録すべき音節
を決定するように成したことを特徴とする音声
登録方式。

a. 発明の詳細な説明

<技術分野>

本発明は音声入力装置における音声登録方式の
改良に関するものである。

<背景技術>

一般に大語彙の音声認識を行なう場合、従来の
ように単語単位で音声を発声して、その音声を登
録していたのでは膨大な音声を全て発声しなければならず、その労力も相当なものになる。また語
彙数の増大による単語間の誤りも顕著になり実用
的でなくなって来る問題点がある。従って、大語彙の音声認識を考えた場合、音声
を詳細に分析し、できる限り音声の特徴を引き出
す必要がある。そこで音声を単語単位でなく、音
素や音節といったより微小時間の単位に分割して、
その単位で識別を行なうことが提案されている。また、音素や音節は数十から数千程度の種類で
あるので、少ない音声の登録で大語彙の音声認識
ができ、有効であると考えられる。しかしながら、各音素や音節は発声毎のばらつ
き、発声の仕方の違い、前後の音声の影響等のた
めに多種多様な変形を受けるため、音声の登録も
それらの影響を予め考慮しておく必要がある。そのため、ある一つの音素や音節に対して、予
め変形を受けた多数の音声を登録しなくてはなら
なくなる。このことは処理量の増大を招き、登録
すべき音声の数も増大してしまうため、音素や音
節単位で識別を行なう利点が失われる。

一方、各音素や音節は全て同じ出現頻度で現われることがなく、また識別率も全て同じではない。出現頻度が高く識別率の低い音素や音節に対してはより注意深く取り扱い必要があるが、出現頻度の極めて少ないものに対してはより多くの処理を行わなくてもよい。

音素や音節の出現頻度や識別率を考慮して音素や音節の登録または再登録を行なうことにより、全体としての性能を維持したままで、より少ない処理量で識別を行なえる可能性がある。

<目的>

本発明は上記のような観点でなされたものであり、より少ない数の音声の登録で、高い認識性能を維持することが可能な音声登録方式を提供することを目的とするものである。

<実施例>

以下、図面を参照して本発明を詳細に説明する。

第1図は本発明の音声登録方式を実施した音声入力装置の一例を示すブロック構成図である。

第1図において、1は音節音声識別部であり、

た音節を指定するコードや番号を入力する。

一方、上記判定結果処理部2内に設けられた音節出現回数計数手段22及び誤り回数計数手段23によって識別されてメモリ21に記憶された各音節の出現回数のカウント数がインクリメントされると共に、誤識別結果指示手段4によって指示された音節の誤り回数のカウント数がインクリメントされ、その結果が各音節毎の出現回数記憶メモリ24及び誤り回数記憶メモリ25に記憶される。

また判定結果処理部2は該処理部2内の変換手段26によって上記メモリ24及び25に記憶された各音節毎の出現回数計数値及び誤り回数計数値から各音節の出現率と識別率を求めて音節出現率表記憶メモリ51及び音節識別率表記憶メモリ52の内容を修正する。

具体的には、例えばメモリ24に記憶された各音節の出現回数のカウント数の総和を求めてから、この総和で各音節の出現回数のカウント数を割った値を各音節の出現率 a_i (i は音節の種類を表わす番号)とし既に音節出現率表記憶メモリ51

該音節音声識別部1では区切って発声された単音節や連続音声から抽出された音節部の認識を行ない、その音節の識別結果を出力して、判定結果処理部2へ送る。上記音節音声識別部1は音節の標準パターンを格納するメモリ11を有し、入力された検出された音節音声から作成された入力パターンと標準パターンとのパターンマッチングを基本として音節識別を行なう。

判定結果処理部2は上記音節音声識別部1から送られて来る識別結果を記憶する識別結果記憶メモリ21を有し、該メモリ21に記憶された内容にもとずいて、一旦、音節、単語、文節、文章等の単位で音節符号を表示部3へ送る。上記表示部3では適当なフォーマットでこれらの文字列を表示することになる。

オペレータは表示部3に表示された識別結果を見て、どの音節の識別結果が誤ったかを見つけ、誤識別結果指示手段4を構成している手動スイッチ、キーボードあるいはライトペン等の手動の入力操作によって、判定結果処理部2に識別の誤っ

に記憶されている音節 i の出現率 b_i と例えば $(kb_i + a_i)/(k+1)$ なる演算(k は適当な値、例えば1~10のある値)によって得られる値を b_i と置き換えてメモリ51に記憶させるように変換手段26が構成されている。同様に誤り回数に対しても、同様の処理が実行され、音節出現回数の総和で各音節の誤り回数のカウント数を割った値を各音節の誤り率 d_i とし、既に音節識別率表記憶メモリ52に記憶されている音節 i の誤り率 e_i と例えば $(he_i + d_i)/(h+1)$ なる演算(h は適当な値)によって得られる値を e_i と置き換えてメモリ52に記憶させる。

このようにして、最近の頻度情報(出現、誤り)を音節出現率表及び音節識別率表に盛り込んでいくことになる。

以上のようにして各音節の出現率及び誤り率を頻度や度数の形で判定結果処理部2内に記憶しておいて適宜(例えば入力音節数がある回数になった時)出現率と識別率に変換し、処理部2内の頻度や度数はリセットされる。

なお、メモリ51及び52内の記憶内容の修正を各音節の処理が終る毎に行なうようにしてもよい。

5は登録判定部であり、該登録判定部5は上記メモリ51及び52に記憶されている出現率表及び識別率表にもとずいてメモリ11に登録されている標準パターンの中の変更を要するものを判定して、その結果を判定結果処理部2へ入力して、表示部8に再登録すべき音節を表示すると共にメモリ11内の所望の音節に対する標準パターンの書き換えを可能な状態にする。

上記登録判定部5の判定動作は各音節に対する誤り率 e_i がある閾値 E を超えたことを判断して行なわれ、この結果として表示部8に音節 i の文字を表示する。

なお、上記閾値 E は出現率 b_i の値に応じて複数個設定されることが望ましく、例えば出現率 B_1, B_2, B_3 ($B_1 > B_2 > B_3$) とした場合、 $b_1 > B_1$ の音節に対しては $e_1 > E_1, b_1 > b_1 > B_2$ の音節に対しては $e_1 > E_2, B_2 > b_1 > B_3$ の音節に対しては

の音声の登録が必要であるかの情報も含めて登録判定部5より再登録すべき音節情報を判定結果処理部2に送る。

ある音節の出現率が高くても識別率が悪い場合には再登録が急がれる。このような場合、判定結果処理部2はこの音節の文字を表示部8で区別表示させるように指示してもよい。例えばディスプレイではその片隅にその文字を点滅させたり、異なる色で表示させたりするように成せばよい。

以上に述べた実施例は音節単位に識別の誤りを指定できる装置に適用した場合であるが、音節単位の識別で単語や文節等を認識する場合、音節の誤りを指定できないことがある。

例えば、単語認識の場合、単語の認識結果をひらがなやカタカナの文字列で表示するよりも漢字で表示した方が分かりやすい。従って、音節単位で誤りを指定するよりも単語単位で誤りを指定した方が全体の処理効率が高くなることがある。

このような場合の処理方法の一例を以下に示す。

今、「たまがわ」と発声した音声入力の識別結

特開59-157699(3)

$e_1 > E_1$ ($E_1 < E_2 < E_3$) の場合に再登録を指示するように成せば、出現率のより高い音節についてはより低い誤り率の場合にも再登録指示が成されることになる。

オペレータは表示部8に表示される音節文字を確認して登録のやり直しをするか否かを判断することになる。

なお、初期の登録の際には、予め別の話者か標準的なものとして得られている出現率表や識別率表に基づいて各音節の標準パターンの個数を算定することになる。

また連続音声でも文の初めや単語の語頭では音節音声の先端部は無音区間の後に続いて現われる。従って、同じ音節でも発声状況によって標準パターンが大変異なったものとなり、音節の標準パターンとしては音声区間中から抽出したものと、無音区間の後から抽出したものが必要となる。その他の調音結合による効果も考えれば、一つの音節に対して多種の標準パターンが必要である。

再登録の際には、一つの音節のどの発声条件で

果として第2図(a)で示すように音節「た」に対して音節候補「か」、「ば」、「た」が得られたとする。

なお、音節文字の下に示されている数値はその候補の信頼性に関する量を変化したものである。この値は信頼性を表わす量と尤度、類似度、距離、第1候補に対する距離比等を用いることができる。ここでは、入力音節のパターンと音節標準パターンとのユークリッド距離を例に示す。各音節候補の信頼度は距離が大きくなるに従って下がる。

単語候補に対する距離を各音節候補の距離の和で表わし、信頼度の高い順に並べると第2図(b)のようになる。入力される単語を地名に限り、単語辞書(後述)に「かまがわ」や「ばまがわ」がなければ地名単語として第2図(d)のようなものが候補として残ることになる。

音節単位の識別で単語を認識する場合には、第3図に示すような音節列候補作成部27、単語辞書28及び辞書照合部29の機能手段を判定結果処理部2は備えている必要がある。ここで音節音声識別部1からは第2図(a)のように信頼度に関す

る量と同時に音節候補が出力される。音節列候補作成部27は第2図(b)のように信頼度の順に音節列候補を作成して辞書照合部29へ送る。辞書照合部29は音節列候補が単語辞書28に有るかどうかを検出し、なければ除外して、単語辞書に有った単語候補を表示部3へ送る。

文章や文節の認識の場合にも、辞書や照合のみならず複雑な処理を必要とするが、いずれにしても音節候補から文節候補、文章候補を表示部3に出力することになる。

第2図(c)に示すように「たまがわ」と発声したにもかかわらず表示された第1単語候補は「神奈川」となったとすると、この場合第2、第3候補まで表示させて、正しい単語は「玉川」であることを指定すべく手動操作で判定結果処理部2に情報を送れば音節候補(第2図(a)から「た」を「か」に誤ったということが分かり識別の正誤の計数が誤り回数計数手段28で行なわれる。

しかし、実用的見地から、第1単語候補が誤った場合には、すみやかに発声し直した方がよい事

がある。この場合、「神奈川」が誤りであることは分かるが、「か」「な」「が」「わ」のどの音節が誤ったかは分からない。これらの各音節の頻度や正誤の度数を計数からはぶくことが妥当な場合もあるが、計数を要する場合には問題である。このような場合、誤った音節列(今の場合「たまがわ」)を一旦記憶しておき、発声し直して正しいと判定された音節列(今の場合「たまがわ」)と比較すれば、「た」を「か」と誤ったことが分かり頻度や正誤の度数に加算することができる。

連続音声から音節部を検出して音節毎に識別する方法よりも、区切って発声された音節の識別の方が一般に識別率が高いと考えられる。従って、連続音声から音節部を検出して音節毎に識別していく方法で、識別を誤った音節を表示部のカーソルを移動して、例えばカナ文字列の相当する位置にもって行って、その音節のみを発声して識別させることによって修正を行うことも考えられる。このとき音声は区切り発声の音節音声になっているので識別は比較的容易である。この場合、同じ

音節を再び同じ音節と誤れば、自動的に別の音節候補に修正することによって、一つのカナ文字を修正するのに数多くの発声は避けられる。

これらの動作を行うための装置の構成例を第4図に示す。

この第4図に示した装置の構成動作を前述の第2図に示した例を用いて説明する。

表示部3には最初の認識結果「神奈川」が表示されているが、誤りであることをキーボード41で指定して、「金沢」と表示し直し、更に誤りを指定した後「玉川」と表示される。このときの単語項目の辞書はメモリ28に格納されている。音節識別結果の文字列「か」「ま」「が」「わ」はメモリ71に記憶されている。

メモリ72は各音節に対して出現回数と誤った回数を格納している。第2図の例の場合、正回答が「玉川」とであると分った後(次の発声を行うと同時に正しい回答であると判断することもできる)、音節照合部20の照合結果によりメモリ72の中の「ま」「が」「わ」の各音節の出現回数のカウ

ント数をインクリメントし、「か」の音節の誤り回数のカウントをインクリメントする。

このようにしてメモリ72は各単語の認識毎に更新されていく。話者が適当な時期にキーボードから指定すればメモリ72の内容を変換手段26によって処理した後メモリ81及び82に記憶された音節出現率表及び音節識別率表をこの処理された内容を用いて更新することができる(この処理を表の更新と呼ぶ)。

この処理の方法としては例えば、一旦、各音節の出現回数のカウント数の総和を求めてからこの総和で各音節の出現回数のカウント数を割った値を各音節の出現率 a_i (i は音節を表わす番号)とし、既に音節出現率表にある音節 i の値 b_i と例えば $(kb_i + a_i)/(k+1)$ なる演算(k は適当な値、例えば、1~10)によって得られる値を b_i と置き換える。これによって、最新の頻度情報を音節出現率表にもり込んでいくことができる。同様に誤り回数に対しても、このような処理を行うことができる。各音節に対する誤り率(音節識別、

率表に格納されている)を e_i とする。

この e_i がある閾値を超えたことを登録判定部6が判定すると、判定結果処理部2を介して表示部8に音節1の文字を表示する。従って、登録のやり直しをするかどうかを話者(使用者)が判断できるようになる。

各音節の出現率は必ずしも一つにする必要はなく、音節当り標準パターンを5個持つような場合には各標準パターンについてカウンタやメモリをもつように成せばよい。

各音節標準パターンは同様の発声状態で得られたものとは限らず、例えば5個の標準パターンの内2個は区切り発声の音節音声から作られていて、残りの3個は単語音声中の音節部から作られていることもある。この場合、音節出現率表、音節識別率表及びメモリ72は各音節毎に値を記憶するのではなく、各標準パターン毎に値を記憶している。

第2図の例で、第1音節「た」の音声は「か」の3番目の標準パターンとの類似度が最も大きか

る。この出現頻度及び誤り度数を求め、この求められた各音節の出現頻度及び誤り度数に関連して登録あるいは再登録すべき音節を決定するように成しているため、全体としての性能を維持したままで、より少ない処理量で再登録を必要とする音節を効率よく見出して、再登録処理することが可能となる。

4. 図面の簡単な説明

第1図は本発明を実施した音声入力装置の一例を示すブロック図、第2図は音声認識例の説明に供する図、第3図は判定結果処理部2の一例を示すブロック図、第4図は本発明を実施した装置の他の一例を示すブロック図である。

1…音節音声識別部、2…判定結果処理部、22…出現回数計数手段、23…誤り回数計数手段、24…出現回数記憶メモリ、25…誤り回数記憶メモリ、3…表示部、4…識別結果指示手段、51…音節出現率表記憶メモリ、52…音節識別率表記憶メモリ、6…登録判定部。

代理人 弁護士 福士 愛彦(他2名)

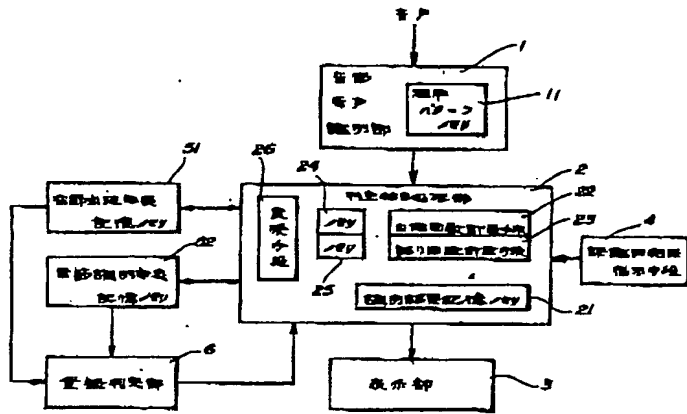
特開昭59-157655(5)

って誤ったとすると「た」の出現回数をインクリメントして「か」の3番目の標準パターンに対応する誤り数のカウンタをインクリメントする。表の更新の結果「か」の3番目の標準パターンに対応する音節識別率表の値 e_i がある閾値を超えた場合にはこの「か」の3番目の標準パターンが作成された音声と同じ発声状態で再登録する必要がある。例えばその音声が「いか」という音声の「か」の部分から作成されたものであれば、表示部に例えば「いか：再登録要」と表示することができる。この場合、音節音声識別部内の標準パターンの各パターンの作成状態を記憶するメモリをメモリ72に付け加えておく必要がある。

以上のようにして再登録の必要性のある音節が音節の出現頻度及び誤り度数にもとずいて表示出力されることになる。

<効果>

以上の如く、本発明によれば、音声入力実行時に得られた音節の識別結果の正誤を指示し、この音節の識別結果及び正誤の指示にもとずいて各音



第1図

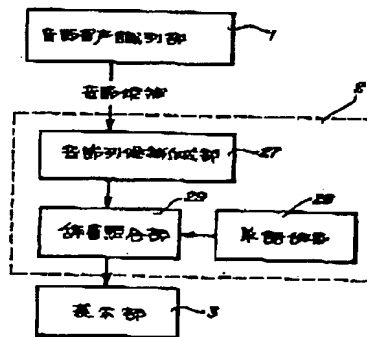
入力音がわ

入力	か	き	く	け
1	1.2	0.8	1.0	1.4
2	1.0	1.1	1.8	2.6
3	1.0		2.4	

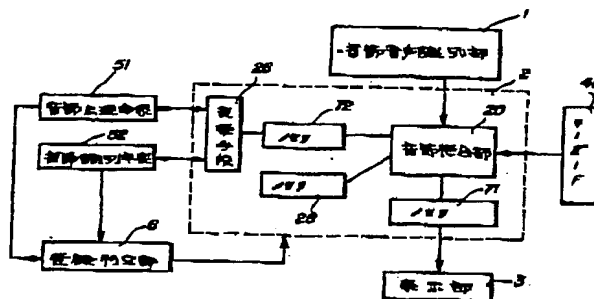
- (b)
1. かきくわ (5.0)
 2. かなくわ (5.5)
 3. ばきくわ (5.4)
 4. かなくわ (5.5)
 5. けきくわ (5.6)

- (c)
1. 海苔川
 2. 金沢
 3. 五川

第2図



第3図



第4図